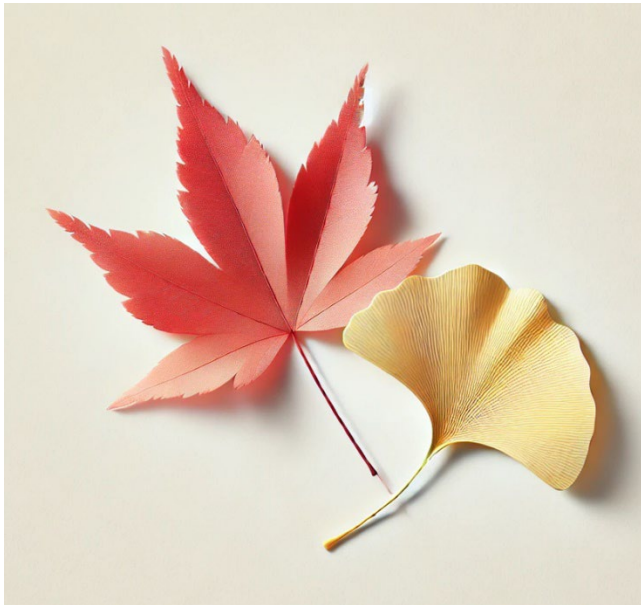


# Assignment 1



BigData Week2

2025. 4.10

Eunhui Kim (김은희)

ehkim@kisti.re.kr

# Assignment 1

- ❖ Submission due : April. 18<sup>th</sup>, 23:59
- ❖ What to submit : Notebook file (.ipynb)
  - Colab : [File]-[Download]-[Download .ipynb]
  - Kaggle : [File]-[Download Notebook]

## ❖ IMPORTANT

- Make sure your results are the same as the output presented on this slide
  - Exception : problem 1 and 8
- Be sure to download the dataset from Assignment-2
  - The file name is "titanic.csv".



# Assignment 1

## Titanic - Machine Learning from Disaster

Start here! Predict survival on the Titanic and get familiar with ML basics



# Assignment 1

- Titanic dataset includes:
  - Passenger ID
  - Passenger Class (1st, 2nd, or 3rd class)
  - Name
  - Sex
  - Age
  - Sibling/Spouse Aboard (SibSp)
  - Parent/Child Aboard (Parch)
  - Ticket Number
  - Fare
  - Cabin Number
  - Port of Embarkation (C = Cherbourg, Q = Queenstown, S = Southampton)
  - Whether the passenger survived (1 for survived, 0 for did not survive)



# Assignment 1

- Titanic dataset includes:

	A	B	C	D	E	F	G	H	I	J	K	L
1	Passenger	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings, I	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen,	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr	male		0	0	330877	8.4583		Q
8	7	0	1	McCarthy,	male	54	0	0	17463	51.8625	E46	S
9	8	0	3	Palsson, M	male	2	3	1	349909	21.075		S
10	9	1	3	Johnson, M	female	27	0	2	347742	11.1333		S
11	10	1	2	Nasser, M	female	14	1	0	237736	30.0708		C
12	11	1	3	Sandstrom	female	4	1	1	PP 9549	16.7	G6	S
13	12	1	1	Bonnell, M	female	58	0	0	113783	26.55	C103	S



# Assignment 1

## Analyze Titanic Dataset by Pandas

- ① Problem 1: Load the Titanic dataset from file – Using the Titanic dataset (titanic.csv) that is uploaded on the LMS



# Assignment 1

## Analyze Titanic Dataset by Pandas

② Problem 2: Show the first 5 rows of the dataset.

index	PassengerId ▲	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.25	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.925	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.05	NaN	S



# Assignment 1

## Analyze Titanic Dataset by Pandas

- ③ Problem 3: Print the Passengers' info rows whose id is from 101 to 104.

index	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
100	101	0	3	Petranec, Miss. Matilda	female	28.0	0	0	349245	7.8958	NaN	S
101	102	0	3	Petroff, Mr. Pastcho ("Pentcho")	male	NaN	0	0	349215	7.8958	NaN	S
102	103	0	1	White, Mr. Richard Frasar	male	21.0	0	1	35281	77.2875	D26	S
103	104	0	3	Johansson, Mr. Gustaf Joel	male	33.0	0	0	7540	8.6542	NaN	S





# Assignment 1

## Analyze Titanic Dataset by Pandas

- ④ Problem 4: Select Passenger info columns such as Name, Age and Cabin.

	Name	Age	Cabin
0	Braund, Mr. Owen Harris	22.0	NaN
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	38.0	C85
2	Heikkinen, Miss. Laina	26.0	NaN
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	35.0	C123
4	Allen, Mr. William Henry	35.0	NaN
...	...	...	...
886	Montvila, Rev. Juozas	27.0	NaN
887	Graham, Miss. Margaret Edith	19.0	B42
888	Johnston, Miss. Catherine Helen "Carrie"	NaN	NaN
889	Behr, Mr. Karl Howell	26.0	C148
890	Dooley, Mr. Patrick	32.0	NaN

891 rows x 3 columns



# Assignment 1

## Analyze Titanic Dataset by Pandas

⑤ Problem 5: Select Passengers whose family members (Sibsp:Spouse and Siblings, Parch: Parents and Child) are on the same Titanic. In other words (Spouse!=0 or Parch!=0)

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2833	C85	C
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
8	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27.0	0	2	347742	11.1333	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
874	875	1	2	Abelson, Mrs. Samuel (Hannah Wizosky)	female	28.0	1	0	P/PP 3381	24.0000	NaN	C
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.1583	C50	C
880	881	1	2	Shelley, Mrs. William (Imanita Parrish Hall)	female	25.0	0	1	230433	26.0000	NaN	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S

354 rows x 12 columns



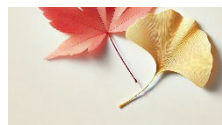
# Assignment 1

## Analyze Titanic Dataset by Pandas

- ⑥ Problem 6: Select the data of passengers who did not survive (Survived == 0)

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S
5	6	0	3	Moran, Mr. James	male	NaN	0	0	330877	8.4583	NaN	Q
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8625	E46	S
7	8	0	3	Palsson, Master. Gosta Leonard	male	2.0	3	1	349909	21.0750	NaN	S
...	...	...	...	...	...	...	...	...	...	...	...	...
884	885	0	3	Sutehall, Mr. Henry Jr	male	25.0	0	0	SOTON/OQ 392076	7.0500	NaN	S
885	886	0	3	Rice, Mrs. William (Margaret Norton)	female	39.0	0	5	382652	29.1250	NaN	Q
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.0000	NaN	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.4500	NaN	S
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.7500	NaN	Q

549 rows × 12 columns



# Assignment 1

## Analyze Titanic Dataset by Pandas

- ⑦ Problem 7: Calculate the average age of passengers who did not survive

30.62617924528302



# Assignment 1

## Analyze Titanic Dataset by Pandas

- ⑧ Problem 8: Show the names of the five passengers who were male, aged 30 or over, traveled in the first class (Pclass==1), and did survive (Survived == 1).
- These are the names of the first 5 people.
  - You can also print the names of another 5 people who meet the conditions.

```
6          McCarthy, Mr. Timothy J
30         Uruchurtu, Don. Manuel E
35   Holverson, Mr. Alexander Oskar
54   Ostby, Mr. Engelhart Cornelius
62      Harris, Mr. Henry Birkhardt
Name: Name, dtype: object
```



# Q & A

김은희 ([ehkim@kisti.re.kr](mailto:ehkim@kisti.re.kr))

TA: Yesim Selcuk ([yesimselcuk@kisti.re.kr](mailto:yesimselcuk@kisti.re.kr))